DOCUMENT RESUME

ED 185 106 TM 800 115

AUTHOR Roberts, Sarah Jane
TITLE Test Floor and Ceiling Effects. ESEA Title I Evaluation and Reporting System.
INSTITUTION RMC Research Corp., Mountain View, Calif.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Jan 78
CONTRACT 300-76-0316
NOTE 23p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Achievement Tests; *Difficulty Level; Educational Testing; Elementary Education; *Grade Equivalent Scores; Guessing (Tests); Mathematical Formulas; Multiple Choice Tests; Standardized Tests; Test Interpretation; *Test Selection
IDENTIFIERS *Ceiling Effects; Elementary Secondary Education Act Title I; *Floor Effects; Out of Level Testing; RMC Models

ABSTRACT

This guide explains the concept of out-of-level testing and suggests a formula for estimating the occurrence of floor effects and ceiling effects, within the context of models for evaluating Elementary Secondary Education Act (ESEA) Title I programs. An analogy explains floor and ceiling effects as if test items are stored in different levels in a warehouse. The impact of guessing on the determination of floor and ceiling levels is also described. Specific suggestions are made for estimating when floor and ceiling effects will occur, and for detecting the presence of such effects from skewness of graphs representing test score distributions. Appended tables provide percentile ranks at which floor and ceiling effects occur for eight elementary-level standardized reading tests: California Achievement Tests (CAT) Comprehensive Tests of Basic Skills (CTBS), Gates-MacGinitie Reading Tests, Iowa Tests of Basic Skills (ITBS), Metropolitan Achievement Tests (MAT), Sequential Tests of Educational Progress (STEP), SRA Assessment Survey, and Stanford Achievement Tests (SAT). Similar tables are provided for seven elementary-level mathematics tests: CAT, CTBS, ITBS, MAT, STEP, SRA, and SAT. Instructions for constructing additional tables and for converting raw score floors and ceilings to percentiles are included. (GDC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TEST FLOOR AND CEILING EFFECTS

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

TM800 115

2 ESEA TITLE 1 EVALUATION & REPORTING SYSTEM

TEST FLOOR AND CEILING EFFECTS

ESEA Title I
Evaluation and Reporting System

Sarah Jane Roberts

January 1978

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

RMC Research Corporation
Mountain View, California

Test Levels

Standardized achievement tests are generally published in several different levels, each containing items with content and difficulty appropriate for children at a particular stage of cognitive development. As a guideline for test users, publishers usually designate each test level as appropriate for certain grades. The Intermediate II level of the Stanford Achievement Test (1973 ed.), for example, is nominally appropriate for children from the end of fifth grade to the end of seventh grade. When children in these grades are tested with this level of the test, they are tested "in level." Strictly speaking, a test level should be designated "in level" only for those grades at which students in the norming sample actually took that level of the test. In other words, a test that is "in level" in spring of third grade should have empirical normative data that were collected at the end of grade three.

It sometimes happens that the lowest achieving students in a particular group will know the answers to only a few items on the nominally appropriate level of a test. In such cases, their scores may reflect chance or guessing as much as true knowledge, and, consequently, will be unreliable. The students have encountered the "floor" of the test. Conversely, the highest achieving students may be able to answer all or nearly all of the items on the nominally appropriate test level. If they could have answered additional, harder items, the test has not given them a chance to demonstrate their true performance levels, and they have encountered the test "ceiling."

An analogy. In order to understand floor and ceiling effects, one can think of the levels of a test as being like the stories of a high-rise warehouse. In the warehouse, test items are arranged on shelves in order of increasing difficulty so

that the easiest item is on the lowest shelf of the basement, and the hardest item is on the highest shelf of the attic. On each level in between, there will be some overlap with adjacent levels; in other words, some of the easiest items will have the same difficulty level as some of the hardest items on the level just below, and some of the hardest items will overlap some of the easiest ones on the level just above.

If a student is sent to a particular level for testing, it is assumed that he could answer all the test items on the lower levels and he is automatically given credit for them. If a student were sent to the third level, for example, his ultimate test score would be the number of third-level items he was able to answer correctly plus the number of items on all lower levels. If we did a good job in selecting the correct level to which to send our hypothetical student, he would find the lower shelves stocked with many items that he could do. As he worked his way up, however, he would eventually reach a point where the items were too difficult. His score, the number of items he completed successfully plus the total number of items on lower levels, would be a reliable measure of his achievement in the area being tested by the items.

If we had done a bad job of estimating the student's ability, we might have sent him to the fourth level where he would find all the items beyond his capability. (We already know that he cannot do the hardest items on the next lower floor.) Unfortunately, the lowest score we could give him would be the sum of all the items below the fourth level--and we know that score would not be a valid index of his true ability. Only if we moved our student down to the next lower level could we assess his true ability because, at that level, he could truly do some of the items. If

we cannot put him on the elevator down, however, the floor of the fourth level represents a lower limit below which his score cannot fall.

Estimating When Floor Effects Will Occur

How do we know when students have reached the floor of a test? We might expect that they would get zero items correct, but that will not necessarily be the case. If all the items are four-choice, students would average about one-fourth of them correct if they simply guessed. Some, however, would be higher and some would be lower. About 10% would get one-third or more of the items correct.

Although few testees resort to purely random guessing, almost everyone will respond to questions when some uncertainty remains regarding the correct answer. For this reason, a score that could easily be obtained by guessing alone is suspect.

In practice, with four-choice items, the floor of the test will begin to influence a significant number of the lowest scores when the group's average score is approximately one-third of the total number of items (Roberts, 1976). In this situation, many of the students will be scoring near zero, and might actually be scoring below that if the floor of the test did not prevent them from doing so. Their scores are thus artificially inflated, as is the group's average score. Experience has shown that the one-third-of-the-items-correct rule of thumb is a reasonably accurate predictor of the threshold below which significant floor effects will be observed for most groups (including Title I students).

Four-choice items are the type found most commonly in achievement tests, but some tests may have three-choice items, or five-choice items, or a combination of several types. The rule of thumb may be generalized to any of these situations. For a

test composed of n-choice items, the floor may be determined using the formula [1/(n - 1) x number of items].

A test composed of several different types of items is simply treated as a group of separate tests combined into one. In other words, the basic formula above is applied to each one and the results are added together.

The rule of thumb should not be regarded as more than that. The likelihood that any particular group will encounter the floor of any particular test will depend not only on the average performance of the group, but on the spread of their true performance levels as well. (The test floor may have been encountered by members of groups with an average performance higher than one-third of the highest possible score and it may not have been encountered by members of other groups scoring lower.)

Estimating When Ceiling Effects Will Occur

At the other end of the scale, "ceiling" effects can artificially depress the scores of students for whom the test is too easy. To return to the analogy of the multi-level warehouse, suppose that our student was sent to the second level, where he found that he was able to do all the items correctly. The second level is too easy a test for him; in fact, he can do many of the items on the third level. Unfortunately, the ceiling of the second level prevents him from showing this ability. Once again, his score will not provide an accurate measure of what he can do.

It is not reasonable to assume that students encounter the ceiling of a test only when they answer all the items correctly, as carelessness, misreading or overinterpreting items, and errors in marking the answer sheet will prevent perfect

scores even when the students could answer additional, more difficult items. As a rule of thumb, we can say that ceiling effects are likely to occur when the average score of a group exceeds three-fourths of the maximum possible score. It should be noted that, unlike floor effects, ceiling effects are not related to the number of choices per item and the theoretical guessing score. Regardless of the number of choices, an average score three-fourths of the total or more will indicate a ceiling effect. This can be more easily seen by considering the score distribution.

Detecting the Presence of Ceiling and Floor Effects

Examining the distribution of scores on a test, one would normally expect it to be roughly symmetrical if the test is of appropriate difficulty for the students. A few students will score very low; the largest proportion of the group will have middle scores, and a few students will score relatively high. This kind of symmetrical distribution will have certain characteristics: The highest and lowest obtained scores will be approximately the same distance above and below the mean, and the mean and the median will be the same.

On the other hand, when the test is not of appropriate difficulty, floor and ceiling effects will show up as skewing of the distribution. In the case of a floor effect, scores will pile up toward the bottom of the distribution because they are being artificially prevented from going any lower. Scores toward the top will be sparse, giving the distribution a long, drawn-out tail at that end, or what is called positive skewing. (Figure 1 depicts a positively skewed distribution.) In the case of a ceiling effect, scores will pile up at the top of the distribution because they are prevented from going any higher. The tail will be long and drawn-out toward the bottom end. This is called negative skewing.

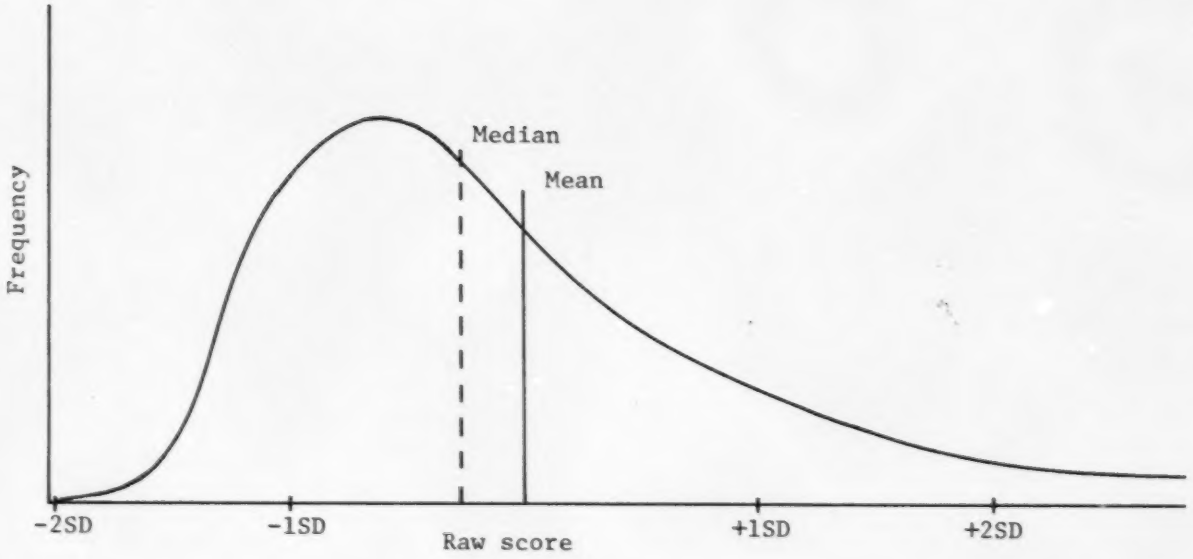

Figure 1. Example of a skewed test score distribution.

These characteristics of the score distribution provide one way of identifying floor and ceiling effects. Ideally, all distributions of test scores used for evaluation purposes should be checked for positive skewing (evidence that the test floor has been encountered) and for negative skewing (evidence that the test ceiling has been encountered) before they are accepted as valid status indicators for the groups tested.

The best method to use in checking for skewness is to plot the frequency distribution of raw scores. Unless the sample size is quite small, visual inspection will usually reveal any significant amount of skewing. A computational alternative is to compare the mean and the median. With smaller samples, random variations are likely to produce some difference and it is not possible to specify an exact point where one should cease attributing the difference to chance and begin suspecting skewness. A difference between the mean and the median of a fifth of a standard deviation or more, however, is almost certain to signal a significant amount of skewing. Checking the distance of the highest and lowest scores from the mean will provide additional evidence.

Ceilings and Floors of Commonly Used Achievement Tests

Tables 1 and 2, presented at the end of this technical paper, summarize information about floor and ceiling effects for eight commonly used, nationally normed tests, assuming in each case that the tests are given in-level. (It should be noted that although both fall and spring testing times are represented in the tables for the sake of allowing comparisons among the tests, not all the tests currently have empirical normative data points in both fall and spring. In Model A evaluations, testing is only recommended at the empirical data points.)

Reading tests. Generally, when the tests are given in-level, a group would have to average at the 10th percentile or lower to encounter the floor, and only a group averaging at about the 70th percentile or higher would be hitting the ceiling. However, there are some notable exceptions to this general pattern that evaluators should know about. For example, the ceiling of the California Achievement Test, Level 2, appears at the 65th percentile at the end of second grade. At the end of the third grade, however, a 36th-percentile group would encounter the ceiling, and at the end of fourth grade a 22nd-percentile group would do so. The test is considered "in-level" for all three grades.

To illustrate how this information can be used to help select an appropriate test level, suppose a local evaluator knew that a Title I group typically had an average performance around the 30th percentile. He might choose to pretest them with the CAT Level 2 at the end of third grade, since the ceiling at that point would only affect groups averaging above the 36th percentile. At the end of fourth grade, however, the evaluator could anticipate that his group would come up against the ceiling on Level 2.

One possible way to avoid this problem would be to use Level 3 at both pretest and posttest time. It would be out-of-level at the end of third grade, but in-level at the end of fourth grade. The danger in this approach would be that the students might find the test too difficult at pretest time. In fact, a check of this hypothesis reveals that a group scoring at the 30th percentile would get substantially fewer than one-third of the items correct on a Level 3 pretest. In this instance the only solution is to give Level 2 as a pretest and Level 3 as a posttest. Although this violates the general recommendation of using the same level pre and post for Model A, and although there is likely to be some change in content between the

two test levels, it would be more satisfactory than using a test where the ceiling or floor effects would prevent the students from showing the gains that might have resulted from the project.

As Table 1 shows, there are a number of cases where the ceiling effect on reading tests comes down as low as the 50th percentile. Here it could affect Model C comparison groups, whose average performance will be above that of the Title I treatment groups. Examples (cited for empirical data points only) are CTBS (1973) Level C at the end of second grade and Level 1 at the end of fourth grade; Gates-MacGinitie (1964) Level D at the end of sixth grade; MAT (1970) Primary I at the beginning of second grade, Primary II at the beginning of third grade, and Elementary at the end of fourth grade; SAT (1973) Primary I at the beginning of second grade, and Primary II at the beginning of third grade. In some cases it would be possible to move to another level normed for the same grade and time, but again one must weigh the possibility of encountering a floor effect.

Floor effects will seldom appear at levels high enough to cause problems with Title I treatment groups when the tests are given in-level. There are exceptions, however. CTBS Level 1 given at the end of second grade will produce a floor effect up as high as the 44th percentile. (ITBS <u>appears</u> to have floor effects rising quite high at all levels, but this is a function of the design of this series, which is <u>not</u> intended to have a wide latitude like the other tests included here. Instead, each level is intended to cover a narrow band of performance and content, and the test is composed of continuous overlapping levels. The publisher recommends that the teacher choose the appropriate level for each child individually, and that different levels of the test be used whenever there are different achievement levels in a group.)

Other instances where floor effects could cause problems occur in the Gates-MacGinitie at the beginning of second and third grades; in the MAT at the beginning of fifth grade; in the STEP (1969) at the end of third and sixth grades; and in the SRA (1971) at the end of grades two, three, four, and five. In these cases, groups averaging above the 20th percentile will encounter the test floor, and the 20th percentile is a likely average for some Title I groups.

Mathematics tests. Table 2 presents floor and ceiling information for mathematics tests. As in reading, the CAT Level 2 shows a descending ceiling. It is safely high (82nd percentile) at the end of grade two, but drops to the 35th percentile at the end of grade three, and to the 14th percentile at the end of grade four. Again these are all officially in-level applications. Other potentially worrisome ceiling effects occur at Level 1 of the CTBS given at the end of fourth grade (37th-percentile ceiling), Level 3 of the CAT given at the end of sixth grade (48th-percentile ceiling), and Primary I of the SRA given at the end of second grade (40th-percentile ceiling). As in the case of reading tests, however, there are often other levels that were normed at the same time and could be appropriate for students who find these levels too easy--provided they do not encounter the floor at pretest time.

The ITBS seemingly shows floor effects at all levels--but see the earlier note on this test. Problems with the test floor are likely to occur with any group below the 30th percentile on the CTBS Level 1 given at the end of second grade, STEP Level 4 at the end of third grade and Level 3 at the end of sixth grade, and the SRA Blue level at the end of third grade.

Tables 1 and 2 contain another column that presents the average number of items that can be answered correctly by children at the 20th percentile. On the Stanford Achievement Test in reading, for example, a raw score of 78 corresponds to the 20th percentile at the end of second grade. This means that, on the average, there are 78 items that can be be done by students at the 20th percentile, as compared with 34, 32, 16, 35, and 20 for the other five tests with information available at that level.

At different levels and for different subject areas, some of the other tests will be found to offer more items in the low range. It is the absolute number of items that concerns us here, rather than what percentage they are of the total. When there are more items at an appropriate level of difficulty for the students in a group, the test is likely to measure their achievement more reliably. Also, the students are less likely to find the test either frustrating or boring.

An evaluator who has a fairly good idea of the general performance level of a local Title I group can use the information in these tables to identify the tests that will measure reliably within the achievement range of the group. Needless to say, it is essential that the test be a valid measure of what has been taught, as well as a reliable measure of student performance, and this validity must be determined by looking at the test content.

Constructing Additional Tables

Some evaluators will wish to construct additional tables similar to those presented here, either because they wish to use some other test or because they wish to use a test level other than the one recommended by the publisher for a given age group. The easiest way to produce such tables is to obtain copies of both the test booklet and

the publisher's norms tables. The first step is then to determine the total number of items. Three-quarters of that number will be the raw score at which the test ceiling occurs, regardless of the number of choices per item.

If all items on the test have the same number of choices, the raw score corresponding to the test floor is determined by dividing the total item count by one less than that number of choices. If the test includes items with different numbers of choices, separate floors must be determined for each item type. The floor of the total test is then the sum of the floors of the different types.

Conversions

Once the raw score ceiling and floor thresholds have been determined, these values can be converted to percentiles using the norms tables. Seperate tables must be used for each grade level of interest. For out-of-level grades, it will be necessary to convert the out-of-level raw score to an in-level percentile. The technical papers entitled Score Conversions and Out-of-Level Testing describe the steps involved in this conversion.

SUMMARY

To summarize, a rule-of-thumb procedure can be used to determine where floor and ceiling effects are likely to occur on any test. Floor effects will occur at or below scores of [1/(n - 1) x number of items], where n is the number of choices per item. Because the test floor prevents the scores of students in the bottom end of the distribution from dropping as low as they really should, pretest scores can appear artificially high, and gains resulting from special treatment will be underestimated.

Ceiling effects will occur at or above scores that correspond to three-fourths of the total items correct. Because the ceiling prevents the scores of students at the top of the distribution from being as high as they really should, post-test scores can appear artificially low, and again gains resulting from special treatment will be underestimated.

REFERENCE

Roberts, A. O. H., Out-of-Level Testing, ESEA Title I Evaluation and Reporting System, Technical Paper. Mountain View, Calif: RMC Research Corporation; January 1978.

TABLE 1

Percentile Ranks Below and Above which Floor and Ceiling Effects Appear for Eight Standardized Reading Tests, Grades 2-6, Fall and Spring

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 20th %ile
CAT (1970 ed.)					
Form A, Level 2	2.0-2.2[1]	85	24	89	26
	2.7		11	65	34
	3.0-3.2[1]		7	53	38
	3.7		4	36	47
	4.0-4.2[1]		3	30	51
	4.7		2	22	62
Level 3	4.0-4.2[1]	82	26	92	24
	4.7		18	85	28
	5.0-5.2[1]		14	79	31
	5.7		9	67	36
	6.0-6.2[1]		8	62	37
	6.7		6	52	41
CTBS (1973 ed.)					
Form S, Level C	2.0-2.2[1]	74	17	65	26
	2.7		10	45	32
Level 1	2.0-2.2[1]	85	63	99	16
	2.7		44	92	20
	3.0-3.2[1]		31	85	23
	3.7		19	71	29
	4.0-4.2[1]		14	61	33
	4.7		9	49	39
Level 2	4.0-4.2[1]	85	31	91	24
	4.7		22	83	27
	5.0-5.2[1]		18	78	29
	5.7		12	67	33
	6.0-6.2[1]		10	62	36
	6.7		9	55	37

[1] Projected, not empirical norms.

TABLE 1 (page 2)

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 20th %ile
GATES[2] (1964 ed.)					
Form 1 & 2 M, Level B	2.1	34	42	90	7
	2.8		8	66	16
Level C	3.1	48	27	84	14
	3.8		12	66	2
Forms 1,2&3M, Level D	4.1	52	18	88	14
	4.8		8	76	19
	5.1		8	69	20
	5.8		4	58	25
	6.1		5	50	26
	6.8		3	42	31
ITBS[3] (1971 ed.)					
Form 5, Level 7	2.2				
	2.7-2.9[1]				
Level 8	3.2				
	3.7-3.9[1]				
Level 9	3.2	60	53	96	13
	3.7-3.9[1]		35	88	17
Level 10	4.2	68	45	96	17
	4.7-4.9[1]		29	88	20
Level 11	5.2	74	33	93	20
	5.7-5.9[1]		20	85	25
Level 12	6.2	76	39	96	20
	6.7-6.9[1]		27	90	23
MAT (1970 ed.)					
Form F, Primary I	2.1	77	12	50	39
Primary II	2.7	84	20	56	35
	3.1		12	38	42
Elementary	3.7	95	20	78	32
	4.1		14	64	38
	4.7		7	50	49
Intermediate	5.1	95	24	84	29
	5.7		14	76	37
	6.1		12	70	38
	6.7		7	62	45

[1] Projected, not empirical norms.

[2] Reading Comprehension only.

[3] The terms "floor" and "ceiling" are not appropriate for ITBS. See text.

TABLE 1 (page 3)

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 20th %ile
STEP,[2] (1969 ed.)					
Form A, Level 4	3.7[3]	60	28	95	18
	4.7		15	83	22
	5.7		8	68	27
Level 3	6.7	60	21	94	19
SRA (1971 ed.)					
Form E, Primary I	2.1[1]	43	18	70	15
	2.7		6	46	20
Primary II	2.1[1]	52	53	94	11
	2.7		30	87	14
	3.1[1]		23	76	16
	3.7		13	59	21
Blue	3.7	90	43	97	23
	4.1[1]		35	93	24
	4.7		23	85	28
	5.1[1]		19	78	30
	5.7		13	67	36
Green	5.1[1]	90	39	93	23
	5.7		29	88	26
	6.1[1]		24	82	28
	6.7		16	73	32
SAT (1973 ed.)					
Form A, Primary I	2.1	147	18	48	78
Primary II	2.8	158	14	52	78
	3.1		16	48	80
Primary III	3.8	125	8	64	55
	4.1		10	62	55
Intermediate I	4.8	127	16	80	45
	5.1		16	78	46
Intermediate II	5.8	121	20	82	40
	6.1		20	80	41
	6.8		14	70	49

[1] Projected, not empirical norms.

[2] Fall norms are projected, not empirical, and are identical to norms for spring of the previous year.

[3] Grade 3 is the lowest grade level tested.

TABLE 2

Percentile Ranks Below and Above which Floor and Ceiling Effects Appear, for Seven Standardized Math Tests, Grades 2-6, Fall and Spring

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 10th %ile
CAT (1970 ed.)					
Form A, Level 1	2.0-2.2[1]	87	6	54	44
	2.7		2	28	59
Level 2	2.0-2.2[1]	117	8	96	39
	2.7		7	82	52
	3.0-3.2[1]		4	64	60
	3.7		1	35	75
	4.0-4.2[1]		1	25	82
	4.7		1	14	95
Level 3	4.0-4.2[1]	108	10	99	34
	4.7		6	96	40
	5.0-5.2[1]		4	90	45
	5.7		2	73	53
	6.0-6.2[1]		2	63	57
	6.7		1	48	64
Level 4	6.0-6.2[1]	98	17	99	25
	6.7		12	98	28
CTBS (1973 ed.)					
Form S, Level C	2.0-2.2[1]	53	17	83	19
	2.7		7	62	25
Level 1	2.0-2.2[1]	98	62	99	19
	2.7		38	94	26
	3.0-3.2[1]		21	81	32
	3.7		11	58	42
	4.0-4.2[1]		7	49	48
	4.7		5	37	55
Level 2	4.0-4.2[1]	98	26	95	31
	4.7		18	88	34
	5.0-5.2[1]		13	80	38
	5.75		8	67	42
	6.0-6.2[1]		7	59	45
	6.7		6	51	48

[1]Projected, not empirical norms.

TABLE 2 (page 2)

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 20th %ile
ITBS[2] (1971 ed.)					
Form 5, Level 7	2.2				
	2.7-2.9[1]				
Level 8	3.2				
	3.7-3.9[1]				
Level 9	3.2	30	24	86	10
	3.7-3.9[1]		8	99	13
Level 10	4.2	36	36	95	10
	4.7-4.9[1]		19	85	10
Level 11	5.2	42	35	97	12
	5.7 5.9[1]		22	90	14
Level 12	6.2	45	35	95	12
	6.7-6.9[1]		22	87	14
MAT (1970 ed.)					
Form F, Primary I	2.1	62[3]	N.A.	N.A.	33
Primary II	2.7	108	5	76	50
	3.1		3	60	56
Elementary	3.7	115	14	82	44
	4.1		10	74	50
	4.7		2	50	65
Intermediate	5.1	115	30	96	33
	5.7		18	88	40
	6.1		14	80	44
	6.7		8	70	52
STEP[4,5] (1969 ed.)					
Form A, Level 4	3.7[6]	60	35	98	17
	4.7		13	85	23
	5.7		5	59	30
Level 3	6.7	60	31	97	18

[1] Projected, not empirical norms.

[2] Test M-1 only.

[3] Of the 62 items on this test, 33 are free-response and only 29 are four-option multiple-choice. Since the norms tables do not separate the two portions, the method for determining floor and ceiling percentiles is not applicable.

[4] Fall norms are projected, not empirical, and are identical to norms for spring of the previous year.

[5] Math computation only.

[6] Grade 3 is lowest level tested.

TABLE 2 (page 3)

Test, Form, Level	Testing Time	Total No. Items	Floor (%ile)	Ceiling (%ile)	Number Items at 20th %ile
SRA (1971 ed.)					
Form E, Primary I	2.1[1]	53	8	69	23
	2.7		1	40	34
Primary II	2.1[1]	58	47	99	14
	2.7		18	98	19
	3.1[1]		12	84	21
	3.7		3	63	31
Blue	3.7	80	54	99	18
	4.1[1]		41	96	20
	4.7		22	90	26
	5.1[1]		18	82	28
	5.7		11	70	34
Green	5.1[1]	80	43	98	20
	5.7		29	96	24
	6.1[1]		24	90	26
	6.7		16	81	30
SAT (1973 ed.)					
Form A, Primary I	2.1	64	4	64	33
Primary II	2.8	100	6	76	46
	3.1		6	70	48
Primary III	3.8	96	12	78	37
	4.1		14	77	38
Intermediate I	4.8	112	18	80	40
	5.1		14	80	42
Intermediate II	5.8	120	16	84	43
	6.1		14	80	46
	6.8		10	66	53

[1]Projected, not empirical norms.